

How to train radiology residents to diagnose pulmonary embolism using a dedicated MRI protocol

Anna Nordgren Rogberg^{1,2}, Sven Nyrén^{2,3}, Eli Westerlund^{4,5}
and Peter Lindholm^{1,2}

Acta Radiologica Open
6(9) 1–6
© The Foundation Acta Radiologica
2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2058460117734244
journals.sagepub.com/home/arr



Abstract

Background: In recent years, magnetic resonance imaging (MRI) has been suggested as an alternative to computed tomography angiography (CTA) to diagnose pulmonary embolism (PE). In previous studies, only senior radiologists have been evaluated as reviewers.

Purpose: To investigate if radiology residents can be trained to review MRI regarding PE and to determine the learning curve effects.

Material and Methods: Four residents independently went through a training program consisting of 70 participants that had undergone steady-state free precession MRI. The individuals were randomized into ten training sessions. For each exam, the review time and presence or absence of embolus was recorded. After completing each session, the residents received feedback on diagnostic accuracy compared to a consensus reading by two specialists. The residents were also presented with the corresponding CTA.

Results: The review time was nearly halved ($P = 0.0002$) during the training program. Comparing the first three sessions with the last three sessions for all residents, the review time decreased from 5:22 min to 2:51 min. The inter-reader agreement improved for all residents during the training program reaching a clinically acceptable level after seven sessions.

Conclusion: Our study suggests that radiology residents can be trained to independently review MRI investigations regarding PE within a short training program. Similar training programs could be more extensively used as effective teaching method for residents.

Keywords

Embolism/thrombosis, education, magnetic resonance angiography, residents

Date received: 19 June 2017; accepted: 30 August 2017

Introduction

In recent years, magnetic resonance imaging (MRI) has been suggested as an alternative diagnostic option for patients with suspected pulmonary embolism (PE), particularly in patients with contraindications to computed tomography angiography (CTA) (1–4). Several different MRI protocols have been described: combined MRI investigations appear to be most reliable while unenhanced MRI protocols show lower risk of complications from contrast media (3). MRI has not been established yet in clinical practice (5) but shows high sensitivity and specificity on a patient based level, in

¹Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden

²Thoracic Radiology, Imaging and Physiology, Karolinska University Hospital, Stockholm, Sweden

³Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

⁴Department of Clinical Sciences, Karolinska Institutet, Stockholm, Sweden

⁵Division of Medicine, Danderyds Hospital, Stockholm, Sweden

Corresponding author:

Anna Nordgren Rogberg, FO Thoraxradiologi, S3:00 Adm Norrbacka, Karolinska University Hospital, SE-171 76 Stockholm, Sweden.
Email: anna.nordgren-rogberg@sll.se



particular for central PE (3). Among the limitations of MRI have been a high proportion of technically inadequate investigations (3,5,6), mainly due to poor vessel opacification and motion artifacts. We have developed a steady-state free precession (SSFP) protocol, with five repetitive slices in each anatomical position instead of respiratory gating, which shows promising results (7).

In previous studies regarding MRI for detection of PE, it is mostly experienced senior radiologists who have interpreted the exams (1,4,5,8). However, in a clinical setting, radiology residents are often the first to review emergency investigations such as cases of suspected PE (9). We wanted to investigate if it would be possible for residents to review MRI exams regarding PE. A supported self-directed training program consisting of a brief introduction followed by training sessions was developed consisting of MRI exams from a previous study (7).

Measurable data regarding the learning process during radiology residency is scarce (10). However, in radiology, training repetitive reviews of similar examinations with instant feedback has shown promising results and appears to stimulate deep learning (11). By plotting learning curves with performance against amount of practice, it is possible to detect at which levels education is most efficient and amount of training required for a certain level of competence (12).

The aim of our study was to investigate if radiology residents can be trained to review MRI regarding PE. We also wanted to determine the learning curve effects including amount of training to be able to review the MRI exams independently.

Material and Methods

Patients

MRI exams from 70 individuals investigated between February 2012 and January 2014 as part of a previously published study by our research group were used to create a training program (7). In the previous study, the diagnostic performance of MRI was compared to CTA. All patients were examined with a diagnostic CTA according to clinical routine, followed by the MRI exam within a maximum of 48 h after the CTA. The local ethics committee has approved the study and informed consent was obtained from each resident participating in the training program within the study. Informed consent from the patients was obtained within the previous study.

CT protocol

The CT exams were performed by a 64-section CT scanner (Lightspeed VCT, GE Healthcare,

Milwaukee, WI, USA) according to the standard protocol for investigating PE at the radiology department. More detailed information regarding parameters can be obtained from our previous study (7).

MRI protocol

The MRI exams were performed on a 1.5-T MRI scanner (Magnetom Aera, Siemens Medical Systems, Erlangen, Germany). The protocol was based on a two-dimensional free-breathing SSFP protocol performed in three orthogonal planes. No intravenous contrast agents or respiratory gating was used. More detailed information regarding MRI parameters can be found in our previous study (7). In each anatomic position five repetitive slices were acquired, which were sorted by position producing stacks with multiple images in different phases of the cardiac and breathing cycle. The total acquisition time was 9:34 min.

Image analysis

Before image analysis, all the MRI exams were blinded and stored in a separate folder in PACS (Sectra Medical Systems, Stockholm, Sweden). Location of PE was assessed based on vascular territories as described by Joshi: central; lobar (right upper lobe, middle lobe, right lower lobe, left upper lobe, lingula and left lower lobe); segmental; and subsegmental (9). Segmental and subsegmental arteries were sorted by their supplying lobar artery.

The residents were instructed to only read and report findings regarding PE in the pulmonary arteries. Secondary findings of PE, such as pulmonary infarction and other potential findings, e.g. pleural effusion, infiltrates, and tumors, were not assessed. The most proximal embolus in each vascular territory was reported. Review time for each examination was also registered.

Training program

The 70 individuals were divided into ten training sessions with seven MRI exams in each. After completing a training session, the resident submitted the results and received feedback on each case compared to a reference standard based on a consensus reading by two specialists. The residents also obtained corresponding CTA exams for comparison before continuing with the following training session.

Four residents in radiology (R1–R4) performed the training program independently after receiving brief instructions. Resident R1 had 3.5 years of experience in radiology and five weeks of experience in general MRI. R2 had three years of experience in radiology and three weeks of MRI experience. R3 had 4.5 years

of training in radiology and three months of practice in MRI. R4 had 2.5 years of training in radiology and two weeks of practice in MRI. Thus, all residents were familiar with CTA to diagnose PE but had limited experience in MRI and none had any previous experience in vascular MRI. To compensate for potential differences in difficulty in the different sessions, two residents (R1 and R2) analyzed the data in chronological order starting with session 1, while the other two residents (R3 and R4) analyzed the data in the reversed order starting with session 10.

Statistical analysis

Regarding review time, descriptive statistics, including mean and range for each reviewer, and MRI protocol were obtained. Significance level P values and correlation r -values were derived by regression analysis, using Excel Data Analysis Tool Pak (Microsoft Office, Redmond, WA, USA).

The outcome of diagnostic accuracy was the agreement between each resident and the reference standard. Kappa values were calculated to determine the agreement for each resident using GraphPad software (13). Due to the small size of each session the sessions were grouped into pairs in chronological order (i.e. 14 patients in each pair of sessions instead of seven in each session). For each resident, the pair of sessions and MRI protocols were analyzed individually.

Results

Diagnostic accuracy

During the training program two residents (R3 and R4) showed a clear improvement in kappa values over time, while the others (R1 and R2) showed a weaker improvement over time (Fig. 1). The improvement over time was primarily due to a large reduction of false positives, from 13 during the first three sessions to two during the last three sessions for all residents together. There was also a small reduction of false negatives, from two during the first three sessions to none in the last three sessions. It was also noted that all residents except for R3 found session 6 more difficult than previous sessions.

After seven training sessions or approximately 50 cases, all residents showed very good or perfect inter-reader agreement compared to the reference for the remaining cases (Table 1).

Review time

The mean review time per examination was 3:56 min (range = 00:13–12:00). The mean review time

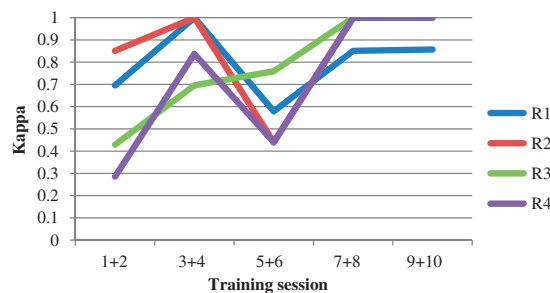


Fig. 1. The inter reader agreement for each pair of session during the training program. Number of sessions in the order the training program was performed (R1 and R2 chronologically and R3 and R4 reversed order).

throughout the training program varied among the residents from 3:04 min to 6:06 min (Table 2), but the pattern over time was similar for three out of four reviewers (Fig. 2). R1 and R2 both showed a steep decrease in time during the first three training sessions (sessions 1–3) and R4 during the first four sessions (sessions 10–7), while R3 showed a gradual decrease in time throughout the training program (sessions 10–1). The decrease in reading time was statistically significant ($P=0.0002$).

The average review time for all exams in the training program was 4 h 35 min (range = 3 h 28 min–7 h 7 min) spread over 1–2 weeks. Average time spent on the training program was estimated to be about one day per resident.

Comparing the mean review time of the first three sessions with the last three sessions for all residents, the mean time was reduced from 5:22 min to 2:51 min. The improvement of reading time during the training program was statistically significant with p -values ranging from 0.001–0.045 (Table 4).

It should be noted that each resident except for R2 had forgotten to record the review time for one exam: R1 in session 3; R3 in session 1; and R4 in session 8.

Distribution of PE

Among the 70 examinations were 18 patients with central PE, three cases of lobar PE, six segmental PE, two subsegmental PE, and 41 cases without PE (Table 3).

Discussion

The primary findings in our study were that residents can be trained to independently review MRI examinations regarding PE within a short training program. The study showed a significant reduction in reading time and improved inter-reader agreement compared to the reference standard during the training program.

Comparing the first three training sessions with the last three sessions, the review time was nearly halved

Table 1. Kappa values for each resident (R1–R4) during the entire training program (sessions 1–10) and for the first (sessions 1–3) and last (sessions 8–10) three sessions.

Sessions	R1	R2	R3	R4
1–10	0.851 (0.725–0.976)	0.907 (0.805–1.000)	0.763 (0.609–0.916)	0.759 (0.602–0.916)
1–3	0.807 (0.558–1.000)	0.904 (0.722–1.000)	0.483 (0.084–0.791)	0.438 (0.084–0.791)
8–10	0.905 (0.724–1.000)	1.000 (1.000–1.000)	1.000 (1.000–1.000)	1.000 (1.000–1.000)

Values are presented as kappa values (95% confidence interval).

Table 2. Mean review time (min) for each resident during the entire training program (sessions 1–10) and for the first (sessions 1–3) and last (sessions 8–10) three sessions.

Sessions	R1	R2	R3	R4	Mean
1–10	03:04 (00:13–10:03)	06:06 (01:00–12:00)	03:12 (00:49–06:27)	03:20 (01:19–07:57)	03:56
1–3	04:55 (00:52–10:13)	07:16 (02:30–15:00)	03:54 (01:59–06:27)	05:14 (02:34–07:57)	05:22
8–10	01:49 (00:13–05:04)	05:33 (01:00–08:30)	02:05 (00:49–04:16)	01:57 (01:19–04:09)	02:51

Values are presented as min (range).

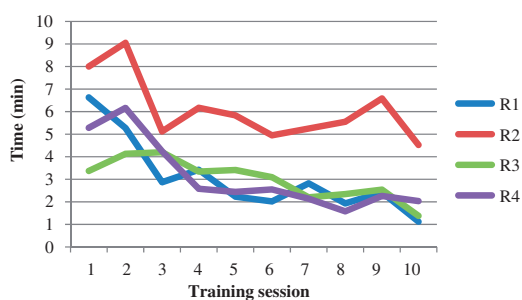


Fig. 2. The average review time for each session during the training program. Number of sessions in the order the training program was performed (R1 and R2 chronologically and R3 and R4 reversed order).

during the training program. Three out of four residents showed a similar learning curve effect, with a steep decrease in review time during the first training sessions (Fig. 2). The pattern was similar regardless of the order of the training program, which favors a learning-curve effect before random differences in difficulty in the different sessions. The resident (R3) with the different learning curve appearance, without a steep decrease during early training sessions, was the most experienced reader and also by far the fastest at the beginning. His learning curve showed a gradual slight decline in reading similar to the other residents following the first sessions of steep decrease. The appearance of PE in MRI and CTA are similar and it is possible that R3 had an advantage of being more experienced in CTA compared to the other residents. However, his result regarding inter-reader agreement at the beginning of the program was not better than the others.

Table 3. Shows the distribution of PE in each training session according to the reference standard.

Exam	Central	Lobar	Segmental P	Subsegmental	No PE
1–7	4	0	0	0	3
8–14	1	0	0	0	6
15–21	1	0	0	0	6
22–28	1	1	1	0	4
29–35	1	0	0	1	5
36–42	1	0	1	1	4
43–49	2	1	1	0	3
50–56	4	0	0	0	3
57–63	2	0	2	0	3
64–70	1	1	1	0	4
Total	18	3	6	2	41

Table 4. Illustrates P values and r-values regarding review time during the training program for each reviewer and for all reviewers together (R1–R4).

	R1	R2	R3	R4	R1–R4
P value	0.002	0.045	0.001	0.002	0.0002
r-value	–0.838	–0.643	–0.872	–0.844	–0.555

The shortest review time was merely 13 s, which would seem extremely short from a clinical point of view. However, this was a case of a saddle embolus affecting both the right and left pulmonary artery. Since the resident had been instructed to only record

the most proximal finding in each vascular territory, no further evaluation was required within the study, which explains the short reading time.

Inter-reader agreement compared to the reference standard illustrated by kappa values was chosen as measure of diagnostic accuracy. Two residents (R3 and R4) showed a clear improvement during the training program while the other two residents (R1 and R2) showed a weaker improvement over time both with a difficulty in session six (Fig. 1). Cohen described kappa values in the range of 0.81–1.00 as nearly perfect to perfect agreement. However, in medical sciences it has been argued that kappa values <0.80 should not be accepted (14). By the final three training sessions, the residents provided kappa values in the range of 0.91–1.00 (Table 1), which would seem to be a clinically acceptable level.

The average review time for the training program was 4 h 35 min. We did not register the amount of time used for feedback and comparison with the reference standard, but the estimated average time spent on the training program was about one day per resident spread over 1–2 weeks. The teaching effort was also limited including the gathering of cases and a short introduction for each resident. Therefore, we argue that similar training programs can be easily applied in most radiological departments.

Two previous studies comparing residents' reading of CTA regarding PE with fully trained radiologists in a clinical setting (9,15) did use discrepancy rates instead of inter-reader agreement. Despite the different statistical methods, the results by the end of our training program appear similar, which also suggests that the diagnostic level of our residents was sufficient. Comparing our results with a previous study on learning curve effects in residents regarding a muscular structure on MRI (10), we found a larger variance regarding initial kappa values, but better results by the end of the training program with nearly perfect or perfect agreement. A contributing factor to our results could be that PE has a similar appearance on CTA and MRI, why detection of PE might be easier than identification of a previously unknown muscular structure. However, there are differences in MRI signal compared to CTA attenuation regarding the assessment of PE, which the novice reader must get used to. Also, the artifacts differ in MRI compared to CTA and might be mistaken for embolus by an inexperienced reader. Among our MRI exams there were cases of phase mismatching, since no gating was used. The repetitive series in each anatomical position were supposed to account for these and we assume they did as the reviewers got used to the exams. There are also a number of flow motion artifacts that may be misinterpreted. It should be mentioned that our study did

not focus on the kind of artifacts occurring but rather on the presence or absence of PE.

There were a high proportion of false positives at the beginning of the training program, probably due to misinterpreted artifacts, while only a few false negatives were recorded. In our previous study where two specialists reviewed the MRI exams, there were no false positives (7), which suggests a learning-curve effect in less experienced readers.

Although the results in our study show promising learning curve effects regarding both time and inter-reader agreement, there are a number of weaknesses in the study design. First, the occurrence of PE and the number of larger emboli was higher than expected, which indicates selection bias. This has been discussed in our previous study (7). Second, the number of residents is small, but the consistent findings support the results. Third, the size of each training session was small, only consisting of seven individuals; however, we wanted the residents to receive feedback within short intervals. In a few groups, there were no true positives, thus it was impossible to calculate kappa values. To solve the problem, consecutive sessions were grouped into pairs and kappa values were calculated based on 14 individuals. Fourth, using a consensus reading of the MRI exams by two specialists instead of the CTA exams as the reference standard makes the results dependent on the specialists' level of expertise. We chose this reference standard because it was known that there were two cases in the training program where isolated subsegmental emboli had disappeared between the CTA and MRI exams. It should also be noted that many studies on learning curve effect use a senior experienced radiologist as the reference (9,10,15). Lastly, each resident except for R2 had forgotten to record the review time for one exam: R1 in session 3; R3 in session 1; and R4 in session 8. However, this is not likely to have had a significant effect on the results.

The training program used in our study was designed as a supported self-directed learning program, which has been suggested as a method to stimulate deep learning (10). Reviewer training and experience are considered important factors in achieving a high diagnostic accuracy (16) but the effects of previous radiologic experience on the ability to learn new radiologic methods have not been determined (10). In our study, we found a learning curve effect for all residents, but the most experienced resident did not perform better on diagnostic accuracy by the end of the training program than the less experienced residents.

In conclusion, our study demonstrates that residents can review our MRI protocol regarding PE following a short training program. It also illustrates the benefits of self-directed training programs when residents are introduced to a new kind of examination.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Swedish Heart and Lung Foundation, the Swedish Society of Medicine and Stockholm City Council.

References

1. Ohno Y, Higashino T, Takenaka D, et al. MR angiography with sensitivity encoding (SENSE) for suspected pulmonary embolism: comparison with MDCT and ventilation-perfusion scintigraphy. *Am J Roentgenol* 2004;183: 91–98.
2. Hochegger B, Marchiori E, Irion K, et al. Magnetic resonance of the lung: a step forward in the study of lung disease. *J Bras Pneumol* 2012;38:105–115.
3. Zhou M, Hu Y, Long X, et al. Diagnostic performance of magnetic resonance imaging for acute pulmonary embolism: a systematic review and meta-analysis. *J Thromb Haemost* 2015;13:1623–1634.
4. Pasin L, Zanon M, Moreira J, et al. Magnetic resonance imaging of pulmonary embolism: diagnostic accuracy of unenhanced MR and influence in mortality rates. *Lung* 2017;195:193–199.
5. Revel MP, Sanchez O, Lefort C, et al. Diagnostic accuracy of unenhanced, contrast-enhanced perfusion and angiographic MRI sequences for pulmonary embolism diagnosis: results of independent sequence readings. *Eur Radiol* 2013;23:2374–2382.
6. Stein PD, Chenevert TL, Fowler SE, et al. Gadolinium-enhanced magnetic resonance angiography for pulmonary embolism: a multicenter prospective study (PIOPED III). *Ann Intern Med* 2010;152:434–443.
7. Nyren S, Nordgren Rogberg A, et al. Detection of pulmonary embolism using repeated MRI acquisitions without respiratory gating: A preliminary study. *Acta Radiol* 2017;58:272–278.
8. Kalb B, Sharma P, Tigges S, et al. MR imaging of pulmonary embolism: diagnostic accuracy of contrast-enhanced 3D MR pulmonary angiography, contrast-enhanced low-flip angle 3D GRE, and nonenhanced free-induction FISP sequences. *Radiology* 2012;263: 271–278.
9. Joshi R, Wu K, Kaicker J, et al. Reliability of on-call radiology residents' interpretation of 64-slice CT pulmonary angiography for the detection of pulmonary embolism. *Acta Radiol* 2014;55:682–690.
10. Tureli D, Altas H, Cengic I, et al. Utility of interobserver agreement statistics in establishing radiology resident learning curves during self-directed radiologic anatomy training. *Acad Radiol* 2015;22:1236–1241.
11. Findlater GS, Kristmundsdottir F, Parson SH, et al. Development of a supported self-directed learning approach for anatomy education. *Anat Sci Educ* 2012; 5:114–121.
12. Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. *Acad Med* 2011; 86:731–736.
13. Software G. GraphPad Software. 2017. Available at: <http://graphpad.com/quickcalcs/kappa1/> (accessed 5 January 2017).
14. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–282.
15. Cervini P, Bell CM, Roberts HC, et al. Radiology resident interpretation of on-call CT pulmonary angiograms. *Acad Radiol* 2008;15:556–562.
16. Dachman AH, Kelly KB, Zintsmaster MP, et al. Formative valuation of standardized training for CT colonographic image interpretation by novice readers. *Radiology* 2008;249:167–177.